DOCUMENT RESUME

ED 417 219 TM 028 164

AUTHOR Behrman, Edward H.
TITLE A Critical Review of New Jersey's High School Proficiency Test in Reading.
PUB DATE 1998-01-00
NOTE 14p.; Version of a paper presented at the Annual Meeting of the California Educational Research Association (Santa Barbara, CA, November 14, 1997) under the title "Promoting Higher Reading Standards via Mandatory Statewide Testing."
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; Grade 11; *Graduation Requirements; *High School Students; High Schools; Reading Achievement; Reading Comprehension; *Reading Tests; Standardized Tests; State Programs; *Test Use; *Testing Programs
IDENTIFIERS New Jersey; *New Jersey High School Proficiency Test

ABSTRACT

This paper describes New Jersey's High School Proficiency Test (HPST) in reading, one of three tests given to all 11th graders as a state graduation requirement. Each reading passage represents one of four test types (narrative, informational, persuasive/argumentative, and workplace). Each passage is followed by multiple-choice items and at least one open-ended item, to measure literal (on-the-line), inferential (between-the-lines), and applied or critical inferential (beyond-the-lines) comprehension. The author concludes that the test responds favorably to several criticisms leveled against standardized reading tests, By using longer text or intact passages drawn from published academic, literary, or institutional sources, the test reasonably reproduces the kinds of academic reading tasks regularly faced by students. However, by failing to measure content-specific comprehension separately, the test does not account for students' differential content schemata. (Contains 19 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 417 219

A CRITICAL REVIEW OF NEW JERSEY'S HIGH SCHOOL PROFICIENCY TEST IN READING

Edward H. Behrman
National University
School of Education and Human Services
9920 S. La Cienega Blvd.
Inglewood, CA 90301

January, 1998

(An earlier version of this paper was presented at the annual meeting of the California Educational Research Association, Santa Barbara, CA, November 14, 1997, under the title "Promoting Higher Reading Standards Via Mandatory Statewide Testing.")

TM028164

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
Edward Behrman
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

ABSTRACT

This paper describes New Jersey's High School Proficiency Test (HSPT) in reading, one of three tests given to all eleventh graders as a state graduation requirement. Each reading passage represents one of four text types (narrative, informational, persuasive/argumentative, and workplace). Each passage is followed by multiple-choice items and at least one open-ended item, to measure literal (on-the- lines), inferentiai (between-the-lines), and applied or critical inferential (beyond-the-lines) comprehension. The author concludes that the test responds favorably to several criticisms leveled against standardized reading tests. By using longer text or intact passages drawn from published academic, literary, or institutional sources, the test reasonably reproduces the kinds of academic reading tasks regularly faced by students. However, by failing to separately measure content-specific comprehension, the test does not account for students' differential content schemata.

A CRITICAL REVIEW OF NEW JERSEY'S HIGH SCHOOL PROFICIENCY TEST IN READING

Edward H. Behrman
National University

The recent mass movement toward statewide standardized testing appears consistent with the charge of the National Commission on Excellence in Education back in 1983 that "standardized testing should be administered at major transition points from one level of school to another" (p.18). Bond & Roeber (1995) reported that all but two states (Iowa and Wyoming) have a statewide assessment in place or in development. Large-scale assessment is used for various purposes, including (1) comparing quality of schools and school districts, (2) measuring individual students' educational progress, and (3) stimulating educational reform.

Many states rely on commercially developed standardized reading tests as part of the assessment battery (Afflerbach, 1990). For example, Nevada and Arizona have used the Comprehensive Test of Basic Skills (Klein, 1995; Statewide Report, 1995) and New Mexico has used the Iowa Test of Basic Skills (Statewide Articulated Assessment, 1995). However, critics argue that many standardized tests do not adequately assess reading competence in light of current reading theory and pedagogical approaches to reading that emphasize student construction of knowledge, reliance on background knowledge or schemata, and metacognition (Farr & Carey, 1986; Levande, 1993). Harker (1990), for instance, states that "standardized reading tests remain locked in a concept of reading which does not coincide with current knowledge of the reading process" and measure an "artificially fragmented and contrived construct of the reading process rather than the highly integrated interactive one which research repeatedly reveals reading to be" (p. 311). Valencia and Pearson (1987) elaborated on the discrepancy between new views of reading and how reading is measured. They noted that most reading tests

1. limit use of prior knowledge by requiring reading of short passages on many topics.
2. lack structural or topical integrity of a large text.

BEST COPY AVAILABLE

3. test literal comprehension rather than inference.
4. rely on multiple-choice items that disregard potential for distractor being a plausible correct answer, based on reader's inference.
5. do not assess readers' strategic approaches to text.
6. do not require reader to synthesize information from various parts of the text.
7. do not assess how well the reader asks (and answers) good questions about text.
8. do not assess reading habits and attitudes toward reading.
9. fragment reading and report scores of isolated skills.
10. do not assess fluency.
11. fail to assess application of knowledge.

Perhaps in response to such criticism, a growing number of states, forsaking reliance on commercially developed reading tests, have developed their own assessment measures (Afflerbach, 1990). One such state, New Jersey, works collaboratively with an outside vendor to produce ongoing editions of its statewide high school graduation test called the HSPT (High School Proficiency Test), which measures students in reading, writing, and mathematics. This paper will first describe the HSPT reading test and then discuss how well it has met the objections raised by Valencia and Pearson that standardized reading tests do not reflect current thinking about the reading process.

In 1995-96 the author spent seven days as a member of New Jersey's reading content committee, reviewing passages and items for inclusion on upcoming reading tests. Findings presented are based on this participant-observer experience, as well as the New Jersey Department of Education's *Report of the Reading Committee* (1990), *Cycle I District Guidelines* (1995), and the *Reading Instructional Guide* (1997).

DESCRIPTION OF THE TEST

Background. In 1988 New Jersey passed legislation moving its High School Proficiency Test (HSPT) from ninth grade to eleventh grade and mandated that, as a graduation requirement, all students entering high school on or after September 1, 1991

must pass the HSPT (New Jersey, 1995). Given to all eleventh graders except LEP (limited English proficient) and special education students, the HSPT reading test is one of three tests (reading, writing, mathematics) that must be passed. All eighth graders are given an Early Warning Test (EWT) to identify students at risk of failing the HSPT; school districts may develop local remedial or other instructional intervention programs to assist these at-risk students. Students who fail the HSPT may retake it later in the eleventh grade and again in the twelfth grade. Students in twelfth grade who meet all other state and local graduation standards but still have not passed all or part of the HSPT must be provided an alternative, untimed assessment (called SRA, Special Review Assessment) to exhibit mastery of HSPT competencies.

Test Development Process. The State Department of Education contracts with an outside vendor to first identify suitable reading passages and then draft test items in strict compliance with test standards developed by the Department (New Jersey, 1990). New passages and items are brought before a reading content committee, comprised of language arts specialists and business persons, to determine if passages and items adhere to test specifications. The committee neither rewrites nor edits items, but accepts or rejects with explanatory reasons. A representative from the external contractor sits in and takes extensive notes during the committee review. Passages and items must also be approved by a sensitivity review committee, comprised of administrators, teachers, and community representatives, to identify potential bias in tasks, statements, or situations presented. Passages and items accepted by both committees are then field tested for future use. Field test items do not count toward a student's score. After field testing, the Department collects data on each test item. The reading content committee then determines that items meet the state's standards for appropriate measurement characteristics. Once approved, the passages and items enter the pool of usable test questions.

Text Type. Each edition of the EWT and HSPT contains reading passages from each of four text types:

1) *Narrative Text:* text that tells a story
2) *Informational Text:* text that conveys information

BEST COPY AVAILABLE

3) *Persuasive/Argumentative Text:* text that is written primarily to convince readers to adopt the writer's point of view

4) *Everyday (EWT) and Workplace (HSPT) Text:* text that people encounter in work and in life. The major difference between everyday and workplace is that grade 8 text is usually school-related (school rules or procedures), while grade 11 text is usually work-related (job application, job search, insurance forms, workplace procedures).

After each text passage are a series of multiple-choice items, including one Knowledge About Reading item, and at least one open-ended item. All passages are expected to be age and grade appropriate, and sensitive to the multicultural nature of the student population. Each passage must possess features germane to specific subcluster skills enumerated for each text type. For example, a narrative text passage must possess strong characterization; influential setting; clear chronological plot; theme on both story-specific and general levels; vocabulary with adequate context clues; and literary devices, such as figurative language. Informational text must possess clear central purpose; major ideas and supporting ideas; structural clues or a visual aid; vocabulary with adequate context clues; and opportunity for future research/study. Persuasive/argumentative text must possess both facts and opinion; main idea and supporting details; an identifiable persuasive technique; use of analogies; vocabulary with adequate context clues; and use of comparison or contrast. Workplace text requires students to decide on an appropriate course of action based on information provided. Workplace text must allow opportunity for synthesizing information; classifying/organizing information; using patterns of sequencing; and extrapolating relevant information.

Text Length and Readability. Each text type must fall within strict parameters of length and readability. Narrative text must contain between 750-2000 words and possess readability in the 9-11 range. Informational text must contain between 700-2000 words and fall into the 8-11 readability range. Persuasive/argumentative text must contain 225-1500 words and possess readability in the 7-10 range, and workplace text must contain 100-500 words and fall into the 9-11 readability range. Readability is based on a readability formula supported by professional judgment.

Source of Text. Narrative text must come from published literary works not previously anthologized; follow a traditional chronological structure with beginning, middle, and end; and have a strong thematic focus. Narrative text may not be adapted but it may be excerpted. Typically narrative passages are high-quality short stories or novel excerpts. Informational text may be selected or adapted from previously published works and must demonstrate clear rhetorical organization reflecting sequence/chronological order, cause/effect, or comparison/contrast. Text must possess two or more levels of information with signals for structure (e.g., subheadings) or visual aid. Informational passages are drawn from textbooks, magazines, or newspapers. Persuasive/argumentative text may be selected or adapted from previously published works and must have a central focusing idea containing facts and opinion. Usually two opposing viewpoints are presented, identified as "Our View" and "Opposing View." Newspaper editorials are a favorite source for persuasive/argumentative text. Workplace text may be selected or adapted from school, government, or business documents. Workplace text may also be addended with hypothetical information about individuals or groups to create simulated decision-making situations.

Task Complexity. The HSPT recognizes that reading comprehension occurs on more than one level; therefore, text passages must contain comprehension items designated as on-the-lines, between-the-lines, and beyond-the-lines. On-the-lines items measure literal comprehension, between-the-lines measure inferential comprehension, and beyond the lines measure applied or critical inferential comprehension. On a typical test, practically all the multiple-choice items are considered either between-the-lines or beyond-the-lines; for example, on the Fall 1994 administration there was only one item identified as on-the-lines (New Jersey, 1997, p. 4-3). The HSPT also recognizes that schema-activation is an important prerequisite to reading comprehension. Therefore, each reading passage includes with directions to students some prereading information or other schema-building activity. Further, because HSPT recognizes that a skilled reader uses a variety of metacognitive strategies to derive meaning, Knowledge About Reading items are used to test the metacognitive process. For example, a student may be asked, "Which of these experiences would most help the reader understand the story? (a) living

BEST COPY AVAILABLE

near woods that are close to a city, (b) knowing someone who was in the army, (c) preparing for a wedding, or (d) caring for someone without him/her knowing" (New Jersey, 1997, p. 7-81). Although scored, Knowledge About Reading items are not included in the overall reading score.

Fixed and Open-Ended Responses. Each of the four text passages is followed by 5-14 multiple-choice items and 1 (sometimes 2) open-ended items. Multiple-choice items retain a traditional format presenting a stem and four answer choices, one the keyed correct choice and the other three distractors. Each multiple-choice item is correlated with a specific subcluster skill. For example, to measure the Using Data Presented in Visual Form subcluster for informational text, a student may be asked, "What is the reason for using the clock graphic? (a) to show the circadian rhythms of the fruit fly, (b) to compare morning rhythms with afternoon rhythms, (c) to illustrate Isaac Edery's personal body clock, or (d) to identify the times of day which are best for specific activities" (New Jersey, 1997, p. 7-105). To measure the Using Patterns of Sequencing to Accomplish a Given Task subcluster for workplace text, a student may be asked, "Pierre is planning on painting his house. To avoid improper disposal of leftover paint, the club recommends that he should *first* (a) carefully calculate the amount of paint needed, and only buy the calculated amount, (b) recycle the empty paint cans, (c) donate the leftover paint to the local theater group, or (d) store the leftover paint for future projects" (New Jersey, 1997, p. 7-135).

Open-ended responses following narrative, informational, and persuasive/argumentative text typically require the student to examine the importance of supporting information. For example, an open-ended item following persuasive/argumentative text may state, "The author of the Our View editorial presents several advantages of turning pro immediately after high school graduation. Identify two or more advantages. Explain why each of these advantages would benefit the athlete. Use information from the Our View editorial to support your answer" (New Jersey, 1997, p. 7-126). Open-ended responses following workplace text typically require the student to develop a course of action and explain its rationale. A sample open-ended item from workplace text reads, "To plan a successful campaign, the East Coast High School

Recycling Committee must make certain preparations for collecting the recyclable materials. Identify two tasks the committee members must do in order to collect the recyclable materials. Explain why each of these tasks is important to the campaign's success. Use information from the text to support your answer" (New Jersey, 1997, p. 7-145).

Scoring Rubrics. The student's response to an open-ended item is graded at score point 3 (beyond-the-lines), 2 (between-the-lines), 1 (on-the-lines), or 0. A generic scoring rubric has been developed to measure reading comprehension on open-ended items. In addition, for each open-ended item actually appearing on a test, a specific scoring rubric must be developed that identifies common elements evident at each score point. The specific rubric is first developed by the outside vendor and reviewed by the content committee. After field testing, the specific rubric is further refined by the content committee through a process called rangefinding. Samples of actual student responses are independently read and scored by committee members, who discuss their scoring and modify the specific rubric to reach consensus for standardized scoring. They create a set of "qualifying papers" identifying model 3, 2, 1, and 0 scores, to assist scorers hired and trained by the vendor.

Reporting Scores. Each multiple-choice item is worth one point and open-ended item three points. A student's total raw score is converted to a scaled score (100 to 500), allowing comparisons across test administrations. A passing scaled score of 300 is required. Although passing is based on total score, score reports break down student performance by text type and level of comprehension.

Influence on Classroom Instruction. From the beginning, HSPT test specifications, sample passages, and items have been available to students, parents, teachers, and administrators (New Jersey, 1990). Because HSPT is a high-stakes graduation test, teachers are highly motivated to help students succeed. In one district, for instance, schoolwide intervention programs and individual classroom instruction incorporate practice on HSPT text types with attention to subcluster skills (Behrman, 1996). In May 1997 the *Reading Instructional Guide* was published "to assist classroom teachers as they continue to link the test specifications and format of the statewide

assessments...to their everyday classroom instructional practices" (New Jersey, 1997, p. 1-2). The testing process is seen as the culmination of classroom activities "to provide rich literacy experiences which will create lifelong readers" (p. 1-3). In coordination with statewide assessment, Core Curriculum Content Standards in Language Arts Literacy were approved by the State Board of Education in May 1996, to be supported by a Curriculum Framework, still under development (New Jersey, 1997, p. 1-2).

DISCUSSION

The HSPT goes a long way toward resolving several flaws of standardized reading tests identified by Valencia and Pearson. Rather than present numerous short passages, the test includes longer passages that allow the reader to draw upon prior knowledge (Issue 1). Although lacking the full structural or topical integrity of a complete textbook or novel, the passages, drawn from published academic, literary, or workplace sources, are sufficiently complete to provide the reader with a realistic approximation of an academic or practical reading assignment (Issue 2). The test places greater emphasis on inferential and critical inferential comprehension than on literal comprehension (Issue 3). Knowledge About Reading items, although not factored into the overall score, assess readers' strategic approaches to text (Issue 5). Particularly on workplace passages, students must synthesize information from several parts of a text and restructure it to both immediate and more distant transfer tasks (Issue 6). The test reports a total reading comprehension score rather than reporting scores of isolated subskills, such as word analysis or vocabulary (Issue 9). Open-ended items, particularly on workplace passages, require students to apply what they have read to a problem-solving application task (Issue 11). Multiple-choice items identified as beyond-the-lines also attempt to measure application of knowledge.

On the other hand, HSPT does not directly respond to four of Valencia and 's issues. Issue 4 (plausible alternative correct answer) is always a concern when oice items are used. To reduce the possibility that a distractor may be a er, the reading content committee rejects such items during the initial

BEST COPY AVAILABLE

review and the item-analysis review. Issue 7 (prediction strategies) is related to metacognition and could be measured as part of the Knowledge About Reading items. Issue 8 (attitude to reading) seems more related to *how* the reader approaches the comprehension activity rather than *how well*, and therefore may not be an appropriate area for summative evaluation. Issue 10 (fluency) sounds like the subskill of word recognition and seems at odds with Valencia and Pearson's position that we should not report performance on these subskills.

Although the HSPT may be considered a laudable first-generation effort at aligning reading assessment with current knowledge about reading, it still must address several underlying weaknesses. By allowing the contractor to excerpt or adapt published works, HSPT may minimally or extensively reduce task authenticity, since students in the "real world" read intact text as it appears in textbooks, novels, magazines, newspapers, or institutional documents. Task authenticity is further reduced by an overly strict adherence to formulaic rhetorical patterns, eliminating the need for students to adjust textual schemata to fit a new rhetorical approach, such as recognizing that a narrative contains an ending-beginning-middle sequence rather than beginning-middle-end. In assigning levels of task complexity to individual items, there is a tendency to overinflate the level of complexity. Often, items whose answers may be found explicitly in the text are classified as inferential (between the lines) rather than literal (on the lines), giving the appearance that the test taps higher-order thinking than it actually does. The open-ended items do create highly authentic tasks but should be expanded, as the balance of multiple-choice items is still extremely heavy: on a typical test, of 48-50 raw score points, only 12-15 points are derived from open-ended items.

Most problematic, however, is HSPT's failure to recognize or account for students' differential content schemata. Reading research, drawn primarily from studies of gender difference, suggests that reading comprehension is highly content-specific (Behrman, 1994). For example, females tend to underperform males on science-related comprehension (American College Testing Program, 1988; Doolittle & Welsh, 1989; Lawrence, Curley, & McHale, 1988; Wendler & Carlton, 1987). While HSPT narrative text is always drawn from a literary source, and workplace text from school, government,

or business documents, there are no parameters for the subject-area sources of informational and persuasive-argumentative passages. On one edition of HSPT, both informational and persuasive-argumentative text may be drawn from civics; on the next, both may be drawn from biological science; on a third, one from economics and one from technology; and so on. Until HSPT controls this variability in content sources of informational and persuasive-argumentative text, it is not known what reading comprehension the test actually measures. At a minimum, the test should measure comprehension of informational text in each of three general content areas (humanities, social studies, and science), and should consistently present the same content area for persuasive-argumentative text.

Despite these limitations, HSPT has had a significant impact on classroom instruction, even in the absence of state content standards. To prepare for the test, students are taught how to read "between the lines" and "beyond the lines" and how to write responses to open-ended items demonstrating critical inferential thinking, in accordance with the generic scoring rubric. While some critics may argue that such "teaching to the test" detracts from other meaningful classroom activities, "teaching to the test" is an appropriate use of class time so long as the test is a valid measure of important educational objectives. To the extent that HSPT presents tasks consistent with real-life academic and workplace reading demands, and fosters higher-order thinking skills, such instructional practices do not seem misplaced.

References

Afflerbach, P. (1990). Statewide reading assessment: A survey of the states. In P. Afflerbach (Ed.), *Issues in statewide reading assessment* (pp. 101-161). Washington, DC: ERIC Clearinghouse on Tests, Measurements, and Evaluation.

American College Testing Program. (1988). *ACT Assessment Program technical manual.* Iowa City, IA: Author.

Behrman, E. (1994, March). *Is reading comprehension global or content-specific?* Presented at the spring conference of the New Jersey Reading Association, Cherry Hill, NJ.

Behrman, E. (1996). *Review of communication curriculum 3-12.* Unpublished: Available from author.

Bond, L. A., and Roeber, E. D. (1995). *The status of state assessment programs in the United States.* (ERIC Document Reproduction Service No. ED 393 859).

Doolittle, A., and Welch, C. (1989). *Gender differences in performance on a college-level assessment test.* (ERIC Document Reproduction Service No. ED 306 237).

Farr, R., and Carey, R. F. (1986). *Reading: What can be measured?* Newark, DE: International Reading Association.

Harker, W. J. (1990). Testing reading: We need a new perspective. *Reading Horizons, 30,* 309-319.

Klein, T. W. (1995). *Nevada Proficiency Examination Program statewide results 1993-94.* (ERIC Document Reproduction Service No. ED 405 351).

Lawrence, I. M., Curley, W. E., & McHale, F. J. (1988). *Differential item functioning for males and females on SAT-verbal reading subscore items.* (ERIC Document Reproduction Service No. ED 304 459).

Levande, D. (1993). Standardized reading tests: Concerns, limitations, and alternatives. *Reading Improvement, 30,* 125-127.

(The) National Commission on Excellence in Education. (1983). *A nation at risk.* Washington, D.C.: U. S. Department of Education.

New Jersey State Department of Education. (1990). *Report of the Reading Committee: Identification of the 8th-grade skills in reading and test specifications and sample items for the 11th-grade High School Proficiency Test and the 8th-grade Early Warning Test.* Trenton, NJ: Author.

New Jersey State Department of Education. (1995). *October 1994 Grade 11 High School Proficiency Test (HSPT11). Cycle I district guidelines: How to interpret and use HSPT11 reports.* Trenton, NJ: Author.

New Jersey State Department of Education. (1997). *Reading instructional guide.* Trenton, NJ: Author.

Statewide articulated assessment system 1994-95 summary report. (1995). (ERIC Document Reproduction Service No. ED 402 354).

Statewide report and appendices for Arizona pupil achievement testing, 1994-95 school year. (1995). (ERIC Document Reproduction Service No. ED 393 869).

Valencia, S., & Pearson, P. D. (1987, April). Reading Assessment: Time for a Change. *The Reading Teacher, 40,* 726-732.

Wendler, C. L., & Carlton, S. C. (1987, April). *An examination of the SAT-Verbal items for differential performance by men and women: An exploratory study.* Presented at the annual meeting of the American Educational Research Association, Washington, DC.

BEST COPY AVAILABLE